RO-SVM: Support Vector Machine with Reject Option for Image Categorization

Rong Zhang and Dimitris Metaxas
Computer Science Deptartment
Rutgers, the State University of New Jersey
Piscataway, NJ 08854 USA

Abstract

When applying Multiple Instance Learning (MIL) for image categorization, an image is treated as a bag containing a number of instances, each representing a region inside the image. The categorization of this image is determined by the labels of these instances, which are not specified in the training data-set. Hence, these instance labels are needed to be estimated together with the classifier. To improve classification reliability, we propose in this paper a new Support Vector Machine approach by incorporating a reject option, named RO-SVM to determine the instance labels, and the rejection region during the training phase simultaneously. Our approach can also be easily extended to solve multi-class classification problems. Experimental results demonstrate that higher categorization accuracy can be achieved with our RO-SVM method, comparing to approaches that do not exclude uninformative image patches. Our method is able to produce results comparable even with few training samples.

1 Introduction

Classifying images into semantically meaningful categories are very important in many applications such as constructing digital image libraries, online image searching, data mining, and surveillance. Automatic image categorization is challenging, especially when to categorize natural images. To address this problem, numerous methods have been proposed to categorize images based on their color and texture properties [14, 3, 6, 4, 20, 19, 5, 18, 11]. In approaches that are most related to this paper, generally multiple regions are extracted from an image, either of regular rectangular shape [14, 5, 18, 19, 11] or of irregular shape [4, 6, 3, 20]. A feature vector, such as color histograms [5, 18, 19], edge direction histograms [19], and wavelet coefficients [20, 3], is then extracted from each image region. Finally, the test image is labeled according to all the information from individual image regions.

Neglecting the spatial correlation between image segments, we can consider an image consisting many regions as a bag containing multiple instances, each representing one region. Hence, image categorization can be formulated as a multiple instance learning (MIL) problem. Current image categorization approaches using MIL applied either the Diverse Density (DD) approaches [15, 22] or the Support Vector Machine (SVM) approaches [2, 6]. Basically, these approaches assume that there is a hidden label for each

BMVC 2006 doi:10.5244/C.20.123

instance, either *positive* or *negative*. A bag is labeled as *negative* if all its instances are negative, or as *positive* if at least one instance in it is *positive*. The main difference between the DD and SVM is the number of concepts to be learnt. A DD approach finds a single concept point, whereas the SVM based approaches aim at finding a proper decision boundary. Hence, for SVM-based approaches, there is no constraint on the number of concepts to be learnt.

As in many learning problems, these derived decision boundary might suffer from over-fitting. This problem is prominent for image categorization, where image regions from different categories may appear similar. A reject option is commonly adopted to safeguard against classification error. The basic principle behind reject option is that when it is not confident for a classifier to label a certain instance, such an instance should not be given a label. Allowing the reject option besides taking a hard decision (-1 or 1) is of great importance in practice [7, 21]. In this paper, we introduce a new SVM-based approach for multiple instance learning, named RO-SVM, where RO stands for reject option. Here, instances that are the most likely to be misclassified are rejected, and the final label for a bag is determined by the remaining instances. Our goal then is to maximize the soft-margin formulation, jointly over the hidden label variables for instances, linear discriminant functions, and the reject region to improve the classification accuracy.

The organization of this paper is as follows: the multiple instance learning with reject option is described in Section 2. Extension of the method to multiple categories is presented in Section 3, followed by the image feature representation in Section 4. Experimental results and analysis are presented in Section 5. Section 6 contains our conclusion and future work.

2 Theoretical Framework for RO-SVM

Classification refers to the problem of selecting a label ω from a label set Ω so that the label is the most appropriate for the given data in a feature space $\Re^d$. If the data is an instance, or a point x in the feature space $\Re^d$, this problem is solved through the traditional pattern classification. If the data is a bag of instances, or a group of points X in the feature space $\Re^d$, the problem is a multiple instance learning problem. In other words, the goal of the traditional pattern classification is to obtain a decision rule $f : \Re^d \rightarrow \Omega$, which aims to provide the most appropriate label for any point in the feature space $\Re^d$. Similarly, the goal of MIL is to learn $f : \mathscr{R} \rightarrow \Omega$, where $\mathscr{R}$ contains multiple instances in $\Re^d$. Hence, compared with the traditional supervised learning, MIL is a more difficult learning problem in a more general setting.

2.1 Goal of Multiple Instance Learning

Let $\{x_1, \cdots, x_n\}$ be the training instances, which are grouped into m bags $\{B_1, \cdots, B_m\}$. Each bag B_I contains a set of instances specified by index set I, i.e., $B_I = \{x_i : i \in I\}$, and is associated with a label $\omega_I \in \Omega$, where the label set $\Omega = \{-1, 1\}$ in the binary case[1]. Denoting ω_i as the instance label for instance x_i, the label of bag B_I can be determined by:

$$\omega_I = \begin{cases} -1 & \text{if } \sum_{i \in I} \frac{\omega_i + 1}{2} = 0 \\ 1 & \text{if } \sum_{i \in I} \frac{\omega_i + 1}{2} > 0 \end{cases} \tag{1}$$

[1] In this paper, we use lower case i for instance index, and upper case I for bag index

The labels for those instances $\omega_i, i = 1, \cdots, n$ are not readily available, whereas the only available labels in the training data-set are those of bags, i.e., ω_I. Therefore, the main problem needed to be solved in MIL is to construct a classifier to determine the labels for individual instances, given a training set with pairs $(B_I, \omega_I), I = 1, \cdots, m$.

2.2 Encoding Reject Option at Instance Level

In MIL, the labels for instances in the training data are needed to be estimated together with the classification rule. To reduce the probability of misclassification, one common approach is to adopt the reject option, i.e., instead of automatically accepting the outcome of a classifier for all points in the sample space, points that do not offer enough confidence in classification are held back. Chow [7] showed theoretically that the optimal rejection rule is to hold back points whose maximum posterior probabilities are less than a certain threshold, i.e., to reject an instance x if:

$$\max_{k=1,2} p(\omega_k|x) < T, \tag{2}$$

where $T \in [0,1]$ is a predetermined threshold which balances the error-rejection tradeoff.

It can be shown that Chow's criteria is a simplified version of the Bayesian decision rule [21]. However, these rules could not be directly applied to MIl for the following reasons: the correct labels for instances are not available anywhere. Only the label for each bag, i.e., multiple instances are provided. Correspondingly, the risk function can not be defied at instance level.

To incorporate the reject option, we introduce a new type of label, 0, for *rejected* instances, in addition to ± 1, and modify the classification criterion as

$$\omega_I = \begin{cases} -1 & \text{if } \sum_{i \in I} \lfloor \frac{\omega_i + 1}{2} \rfloor = 0 \\ 1 & \text{if } \sum_{i \in I} \lfloor \frac{\omega_i + 1}{2} \rfloor > 0 \end{cases} \tag{3}$$

where the function $\lfloor x \rfloor$ gives the largest integer less than or equal to x. Note here the reject option is encoded at instance level in order to predict bag labels more precisely.

2.3 RO-SVM

We choose to solve our problem using Support Vector Machines, which are widely used in building a binary instance-level classifier [8]. The SVM algorithm determines a hyperplane to divide the training data into two classes with the maximum-margin, i.e., the distance between the hyperplane and the closest training data, known as the margin, is maximized. By applying the kernel trick [1], the SVM algorithm is able to fit the maximum-margin hyperplane in a transformed space induced by the kernel function. This transformation may be non-linear and the transformed space may be high dimensional. Therefore, although the classifier is a linear separator in the high-dimensional transformed space, it may be non-linear in the original feature space. Details of SVM can be found in [8].

Obviously, it is not feasible to directly apply SVM approaches on MIL, because only labels for bags are provided in the training data, not those for the individual instances. If we treat the labels of individual instances as hidden variables that need to be determined together with the hyperplane parameters of SVM at the time of training, we need to solve the following optimization function:

$$\min_{\{\omega_i\}} \min_{\mathbf{w},b,\xi} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_i \xi_i \tag{4}$$

subject to the constraint in Eq. 1 and

$$\omega_i(<\mathbf{w}, x_i> +b) \geq 1 - \xi_i, \xi_i \geq 0, \omega_i \in \Omega, \tag{5}$$

where ω_i is the instance label, $\mathbf{w}$, and b are parameters for the decision boundary, and ξ_i is a slack variable which is introduced to allow some classification errors. The above formulation is a mixed integer programming problem and could be solved through local optimization heuristics [2].

To encode reject option for improving classification confidence, most available methods reject patterns whose distance from the decision boundary is lower than a predefined threshold after the training phase [10]. However, as pointed out by Fumera et. al. [10], the rejection region must be determined during the training phase in order to obey the structural risk minimization principle, on which SVMs are based. For this purpose, in addition to the constraint in Eq. 3 and Eq. 5, we also introduce a slightly different optimization function as:

$$\min_{\{\omega_i\}} \min_{\mathbf{w},b,\xi} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{\xi_i \geq T} \xi_i \tag{6}$$

where $T \geq 0$ is a predefined threshold. In this formulation, errors introduced by instances located close to the decision boundary are not encoded in this optimization function. In fact, this problem cannot be solved efficiently. To estimate the decision rule and the corresponding reject region during the training phase, we propose to alternately update the hyperplane parameter estimation and the instance label estimation during optimization procedure. More specifically, based on the current instance labels, a hyperplane is constructed through SVM using only informative instances, which are instances located away from the decision boundary, i.e. $\omega_i \neq 0$. The resulting hyperplane in turn provides a new label for every instance. When updating new labels for instances, we have to make sure all the additional multiple instance constraints in Eq. 3 are satisfied. The above two steps are processed iteratively until no more changes appear in the instance labels, as detailed in Table 1.

initialize $\omega_i = \omega_I$, for $i \in I$
repeat
    compute SVM solution $\mathbf{w}, b$ for (x_i, w_i) pair with $\omega_i \neq 0$
    compute outputs $f_i =< \mathbf{w}, x_i > +b$ for all x_i
    set $\omega_i = sgn(f_i)$ for every instance i in positive bags
    set $\omega_i = 0$ for all instance i with $|f_i| < T - 1$
    for every positive bag B_I
        if $(\sum_{i\in I}\lfloor(1+\omega_i)/2)\rfloor == 0$
            compute $i^* = \arg\max_{i\in I} f_i$
            set $\omega_{i^*} = 1$
        endif
    endfor
until no change in any ω_i

Table 1: Pseudocode of the RO-SVM optimization.

3 MIL for Multiple Categories

In many cases, it is more desirable to categorize images into multiple categories. We show here that our RO-SVM algorithm can be extended to multiple category cases.

In MIL, positive bags may contain negative samples, while negative bags can not contain any positive samples, suggesting that the positive and negative classes are not equivalent. This is the biggest difficulty in extending MIL to multi-class problem. However, this consideration comes from drug discovery applications when MIL was first introduced [15]. In image categorization, such assumption could be too strong. It is possible that we loosen the MIL assumption of "at least one of the elements in this bag belong to class X" to "most of the elements in this bag belongs to class X". This new assumption would allow a small amount of outliers, which is formulated as:

$$\omega_I = j \;\text{ if } \sum_{i \in I} \lfloor \frac{\omega_i}{j} \rfloor > \alpha \max_{k \neq j} (\sum_{i \in I} \lfloor \frac{\omega_i}{k} \rfloor) \tag{7}$$

where $\alpha > 1$ is a scaling factor controlling the amount of outliers we may allow in a bag. Correspondingly, the procedure of RO-SVM solving multi-class MIL problems is summarized in Table 2. In our implementation, we employed [12] for solving the multi-class SVM problem.

4 Image Feature Vector Representation

Due to the fact that semantically meaningful segmentation is still an open problem in computer vision [23], we choose to use image patches with regular shapes. For each image in the training data-set, we first divide it into several overlapping rectangular image patches. A feature vector is then calculated to represent each image patch. As exact feature representation is not necessary for our purpose, we only need a feature representation that is informative enough to distinguish both color differences in objects such as sky, mountain, grass, etc., and the texture differences for ocean waves, elephant skins, etc. In our experiment, an image patch is represented as a vector containing the spectral histogram with multiple marginal distributions of responses, as suggested in [13].

```
initialize ω_i = ω_I for i ∈ I
repeat
    compute multi-class SVM solution for (x_i, w_i) pair with ω_i ≠ 0
    compute outputs f_ji =< w_j, x_i > +b for all instance x_i, and class ω_j
    set ω_i = argmax_j f_ji for every instance i
    set ω_i = 0 for instance i with max_j f_ji < T − 1
    for every bag B_I with label j
        while (∑_{i∈I} ⌊fracω_i j⌋ < α max_{k≠j}(∑_{i∈I} ⌊ω_i/k⌋)
            compute i* = argmax_{i∈I,ω_i≠j} p(ω_j|x_i)
            set ω_{i*} = j
        endwhile
    endfor
until no change in any ω_i
```

Table 2: Pseudocode of the RO-MIL for multi-class problems.

Suppose we have K filters $\{f^{(\alpha)}\}, \alpha = 1, 2, \cdots, K$. For each filter $\{f^{(\alpha)}\}$, we compute a sub-band image $P^{(\alpha)}$ through linear convolution

$$P^{(\alpha)}(\overrightarrow{v}) = f^{(\alpha)} * P(\overrightarrow{v}) = \sum_{\overrightarrow{u}} f^{(\alpha)}(\overrightarrow{u})P(\overrightarrow{v} - \overrightarrow{u}), \tag{8}$$

at each pixel location $\overrightarrow{v}$. Then we define a histogram as

$$h_P^{(\alpha)}(z) = \sum_{\overrightarrow{v}} \delta(z - P^{(\alpha)}(\overrightarrow{v})), \tag{9}$$

where $\delta(.)$ is the Dirac delta function. Assuming that the filters are independent of each other, we concatenate the spectral histograms to form a feature vector as:

$$h_P = (h_P^{(1)}, h_P^{(2)}, \cdots, h_P^{(K)}). \tag{10}$$

The feature vector representation defined in Eq. 10 is invariant to translation, because the filter responses only depend on the relative locations of pixels. With sufficient number of filters, the above spectral histogram can uniquely represent any image[13].

From visual perception and empirical studies of independent components of natural images [16], four different types of filters are suggested to be used in practice:

- The intensity filter, which is the image value at a given pixel.
- Gradient filters, which capture the edge information.
- Laplacian of Gaussian filters [17].
- The Gabor filters [9].

Here, we choose a total of ten filters: three intensity filters for the three color channels (RGB), two second order gradient filters, two Laplacian of Gaussian filters, and three Gabor filters at different orientations of the same scale. The last three groups of filters are applied to the corresponding gray-scale images. One could further apply dimensionality reduction methods to reduce the dimensionality of this feature representation.

5 Experiments

To evaluate our method in image categorization problems, we applied our method on the WANG data-set [20] [2], a widely used test-bed for image categorization. The data-set consists of 1000 images in JPEG format with size 256×384 or 384×256. There are a total of ten different categories, each containing 100 images. The ten categories (labeled from 0 to 9 successively) are: African people and villages, beaches, buildings, buses, dinosaurs, elephants, flowers, horses, mountains and glaciers, and food.

One important factor that needs to be predefined for our method is the size of image patches. Our spectral histogram image feature representation will fuse the texture information for large patches, making conceptually different image patches similar in representation. On the other hand, smaller patches contain less texture information, which may lead to different representations even for similar patches. In addition, a small patch size could result in a large number of patches for the training step, which may not be computationally feasible. The patch size we chose in our experiments was 51×51, which was tuned experimentally with a small number of testing images. For images with 256 pixels in height and 384 pixels in width, we extract 192 overlapping 51×51 image patches,

[2] This data-set can be downloaded from http://wang.ist.psu.edu/docs/related

	Method	Accuracy (standard deviation)
Patch	RO-SVM	89.4% $\pm$ 1.2%
	mi-SVM	60.1 %(3.5%)
	MI-SVM	63.6 % (3.6 %)
Region	DD-SVM	81.5%(1.5%)
	Hist-SVM	66.7%(1.1%)
	MI-SVM	74.7%(0.3%)

Table 3: Comparisons of classification accuracy of image categorization experiments with different methods. The images belong to Category 0 (African people and villages) and Category 9 (food).

whose centers are placed on the grid of 12 horizontal lines and 16 vertical lines equally spaced inside the images. For images with 384 pixels in height and 256 pixels in width, we transpose the mesh grid, and extract the image patches accordingly.

For all of our experiments, either binary or multi-class, we only use ten labeled images for each class as the training data, following the same setting in [15]. The reason for using such a small number of training data is that in typical image categorization scenarios, there might be only a few training examples available. We repeat each experiment 10 times with random training data, and report here the average of the results obtained.

5.1 Classification Results for Two-Class Problems

First, we employed our RO-SVM method as stated in Table 1 for categorizing images for two categories.

Results provided in Table 3 are categorization accuracy for images in Category 0 (African people and villages) and Category 9 (food)[3]. In the first three approaches, the same image feature representations for image patches as presented in section 4 are used, where the mi-SVM and MI-SVM are approaches that maximize pattern margin and bag margin respectively [2]. Our RO-SVM achieves a significantly higher accuracy with a lower standard deviation, comparing to methods which did not incorporate the reject option. Furthermore, our RO-SVM performs better than region based approaches reported in [6], which used features for image regions obtained from image segmentation. Here, DD-SVM is a DD based SVM approach, and Hist-SVM is a histogram-based SVM approach. Note that in the experiments for the region based approaches, results are obtained using half of the available images in training [6], while we only use 10% of the data. This suggests that our RO-SVM could achieve a higher accuracy even with a smaller number of training data, while at the same time avoiding complicated image segmentation procedures.

Furthermore, we performed binary categorization on other categories with results shown in Table 4. Although in both cases, RO-SVM outperforms mi-SVM and MI-SVM in both accuracy and standard deviation, all approaches get high accuracies when comparing horse images verses elephant images, but perform much worse to separate images of beach scenes from Mountains and glaciers.

The reason for the difference in categorization for different data-set lies in the data itself. To separate a horse from an elephant, color and texture information from the animal

[3] We follow the same experiments as done in [6]

Data Set	RO-SVM	mi-SVM	MI-SVM
Cat. 7 vs Cat. 5	86.7% (1.3 %)	78.6% (3.5%)	81.9% (4.4%)
Cat. 1 vs Cat. 8	71.7% (1.7%)	62.4% (8.1%)	61.0% (4.1%)

Table 4: Comparisons of classification accuracy of image categorization experiments with different data.

Figure 1: First three images: beach scenes which are categorized as mountain view. Last three images: mountain views which are categorized as beach scenes.

region is dominant, while all other environment information can be irrelevant. Therefore, it is possible to label an image as long as one patch in the image appears like the animal, which is exactly the basic MIL assumption. However, for the beach scene and mountain view images, this assumption is not very accurate. Figure 1 shows some mislabeled images in these two categories. The appearances in these images are relatively similar to each other, though human experts could categorize them with strong prior knowledge of these two kinds of views. One possible prior knowledge is that if there exist sand areas in a given image, it would be categorized as a beach scene. However, the color and texture for these sand areas look similar to snow on glaciers. Therefore, these categorization errors may not be avoidable in low level feature based categorization approaches.

5.2 Multi-class Categorization Accuracy

We also tested our multi-class RO-SVM on the entire Wang data-set. For the 1000 images in the WANG data-set, our method yields an overall accuracy of 83.3% with standard deviation 8.9%. In contrast, previously reported accuracy on this data-set vary from 37.5% to 84.1% [14]. Our method is comparative to the most effective image categorization methods available. Further, direct comparison of the accuracy rate is not appropriate, since most of the results are obtained with a large number of training data. For example, Marée et. al. [14] applied leave-one-out cross-validation strategy, which means they trained a classifier using 999 images to classify the remaining single image. In our approach, as stated above, we only use a total of 100 images for training, ten from each category, and test on the remaining 900 images. The result from our experiment shows that our method is very effective even using a quite small number of training data.

In Table 3, we provide the confusion matrix of our image categorization results. Hence, the numbers on the diagonal of the matrix represent the classification accuracy for each category. Off-diagonal elements indicate classification errors.

Two most confusing categories are category 1 (beaches) and category 8 (mountains and glaciers): 10% of beach scenes are mislabeled as mountains and glaciers, whereas 8% mountains and glaciers images are mislabeled as beach scenes. This observation is consistent with that of Chen et al.[6]. Some of the images from the two categories are similar, and it is difficult to really separate them unless prior knowledge is involved.

	C. 0	C. 1	C. 2	C. 3	C. 4	C. 5	C. 6	C. 7	C. 8	C. 9
Cat. 0	78	0	2	0	0	12	1	1	4	1
Cat. 1	0	80	6	0	0	2	2	0	10	0
Cat. 2	1	6	61	13	0	7	3	0	4	4
Cat. 3	0	2	1	92	0	0	0	0	0	4
Cat. 4	0	0	0	0	100	0	0	0	0	0
Cat. 5	2	0	7	0	0	81	0	6	3	1
Cat. 6	0	0	0	0	0	1	98	0	0	1
Cat. 7	2	1	1	0	0	10	1	83	1	0
Cat. 8	0	8	7	1	0	6	0	1	78	0
Cat. 9	3	0	1	8	1	1	0	1	2	82

Table 5: Confusion matrix in percentage of experiments for multiple image categories. Only 10 images per category are used as training data, and the remaining 90 images for each category are used as testing data.

6 Conclusion and Future Work

In this paper, we presented a novel SVM based approach with reject option, or RO-SVM, for image categorization problems, where images are represented as an ensemble of rectangular image patches. The rejection option is necessary in both training and testing procedure to reduce classification error, which may be caused by the ambiguity in determining the labels for those image patches that lie in the vicinity of decision boundary. In addition, we generalized our RO-SVM algorithm to multi-class learning, which is more desirable in general classification scenarios. This framework was applied to automatic image annotation experiments on the WANG data-set. The results from two-class categorization experiments showed that the incorporation of rejection option in SVM significantly improves the classification accuracy. For multi-class problems, our method achieved a high categorization accuracy using only a small number (10%) of the training data.

One important issue in our experiments is the optimum threshold setting in reject option. Currently, we use only a single threshold to reject instances over all classes. As the posterior probability measurement is just an estimation, different thresholds for different classes may be preferred. In addition, further study will be carried out to design new image feature representations that are invariant to rotation and facilitate robust distance measures between image patches.

Acknowledgments

The authors would like to thank National Science Foundation, and DIMACS Graduate Student Awards for supporting this work.

References

[1] M. A. Aizerman, E. M. Braveman, and L. I. Rozoner. Theoretical foundations of the potential function method in pattern recognition learning. *ARC*, 25:821–837, 1964. 3

[2] S. Andrews, I. Tsochantaridis, and T. Hofmann. Support vector machines for multiple-instance learning. In *NIPS*, pages 561–568, 2002. 1, 4, 7

[3] J. Bi, Y. Chen, and J. Wang. A sparse support vector machine approach to region-based image categorization. In *CVPR '05*, 2005. 1

[4] C. Carson, S. Belongie, H. Greenspan, and J. Malik. Blobworld. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(8):1026–1038, 2002. 1

[5] O. Chapelle, P. Haffner, and V. Vapnik. SVMs for histogram-based image classification. *IEEE Transactions on Neural Networks*, 10(5):1055–1064, 1999. 1

[6] Y. Chen and J. Wang. Image categorization by learning and reasoning with regions. *J. Mach. Learn. Res.*, 5:913–939, 2004. 1, 7, 8

[7] C. Chow. On optimum recognition error and reject tradeoff. *Information Theory, IEEE Transactions on*, 16(1):41–46, 1970. 2, 3

[8] N. Cristianini and J. Shawe-Taylor, editors. *AN INTRODUCTION TO SUPPORT VECTOR MACHINES*. Cambridge University Press, UK, 2000. 3

[9] J.G. Daugman. Complete discrete 2d gabor transforms by neural networks for image analysis and compression. *ASSP*, 36(7):1169–1179, 1988. 6

[10] G. Fumera and F. Roli. Support vector machines with embedded reject option. In *SVM*, pages 68–82, 2002. 4

[11] M.M. Gorkani and R.W. Picard. Texture orientation for sorting photos at a glance. In *IEEE International conference on Pattern Recognition*, volume 1, pages 459–464, 1994. 1

[12] T. Huang, C. Lin, and R. C. Weng. A generalized bradley-terry model: From group competition to individual skill. In *Advances in Neural Information Processing Systems 17*, pages 601–608, 2005. 5

[13] X. Liu and D. Wang. Texture classification using spectral histograms. *IEEE Transactions on Image Processing*, 12(6):661–670, 2003. 5, 6

[14] Ra. Marée, P. Geurts, J. H. Piater, and L. Wehenkel. Random subwindows for robust image classification. In *CVPR*, 2005. 1, 8

[15] O. Maron and A. L. Ratan. Multiple-instance learning for natural scene classification. In *ICML*, pages 341–349, 1998. 1, 5, 7

[16] B. A. Olshausen and D. J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–609, 1996. 6

[17] G. Sotak and K. Boyer. The laplacian-of-gaussian kernal. *CVGIP*, 48(2):147–189, 1989. 6

[18] M. Szummer and R. W. Picard. Indoor-outdoor image classification. In *Workshop on Content-based Access of Image and Video Databases*, 1998. 1

[19] A. Vailaya, M.A.T. Figueiredo, A.K. Jain, and H. Zhang. Image classification for content-based indexing. *Image Processing, IEEE Transactions on*, 10(1):117–130, 2001. 1

[20] J. Wang, J. Li, and G. Wiederhold. Simplicity: Semantics-sensitive integrated matching for picture libraries. *PAMI.*, 23(9):947–963, 2001. 1, 6

[21] A. Webb. *Statistical Pattern Recognition*. John Wiley and Sons, UK, 2002. 2, 3

[22] Qi Zhang, Sally A. Goldman, Wei Yu, and Jason Fritts. Content-based image retrieval using multiple-instance learning. In *ICML*, pages 682–689, 2002. 1

[23] Song Chun Zhu and Alan Yuille. Region competition: Unifying snakes, region growing, and bayes/mdl for multiband image segmentation. *PAMI.*, 18(9):884–900, 1996. 5